

Can I Have Your Recipe? Using a Fidelity of Implementation (FOI) Framework to Identify the Key Ingredients of Formative Assessment for Learning

Erika G. Offerdahl,^{1*} Melody McConnell,² and Jeffrey Boyer³

¹School of Molecular Biosciences, Washington State University, Pullman, WA 99164-7520;

²Department of Biological Sciences and ³Dean's Office, College of Science and Mathematics, North Dakota State University, Fargo, ND 58108

ABSTRACT

For decades, formative assessment has been identified as a high-impact instructional practice that positively affects student learning. Education reform documents such as *Vision and Change: A Call to Action* expressly identify frequent, ongoing formative assessment and feedback as a key instructional practice in student-centered learning environments. Historically, effect sizes between 0.4 and 0.7 have been reported for formative assessment experiments. However, more recent meta-analyses have reported much lower effect sizes. It is unclear whether the variability in reported effects is due to formative assessment as an instructional practice in and of itself, differences in how formative assessment was enacted across studies, or other mitigating factors. We propose that application of a fidelity of implementation (FOI) framework to define the critical components of formative assessment will increase the validity of future impact studies. In this *Essay*, we apply core principles from the FOI literature to hypothesize about the critical components of formative assessment as a high-impact instructional practice. In doing so, we begin the iterative process through which further research can develop valid and reliable measures of the FOI of formative assessment. Such measures are necessary to empirically determine when, how, and under what conditions formative assessment supports student learning.

INTRODUCTION

The formative use of assessment has long been identified as critical to teaching for conceptual understanding (e.g., Sadler, 1989; Black and Wiliam, 1998a; Bell and Cowie, 2001). Formative assessment has been widely cited as a high-impact instructional practice; indeed, it has become an “urban legend” as perhaps the single most effective instructional intervention to increase student learning (Briggs *et al.*, 2012). Effect sizes between 0.4 and 0.7 have historically been reported based on work by Black and Wiliam (1998a,b). However, sparse description of the methodology involved in calculating these statistics has led to recent scrutiny of the reported effect (Dunn and Mulvenon, 2009; Kingston and Nash, 2011; Briggs *et al.*, 2012). A more recent meta-analysis calculates a much smaller effect size, on the order of 0.2 (Kingston and Nash, 2011). Further, the methods by which effect sizes were calculated may have biased the estimate (Briggs *et al.*, 2012).

The empirical evidence is variable, but the link between formative assessment and student learning is intuitive and has strong theoretical support (e.g., see Sadler, 1998; Furtak and Ruiz-Primo, 2008; Black and Wiliam, 2009; Bennett, 2011; Offerdahl and Montplaisir, 2014). The literature is immense, which means that formative assessment may assume different names (i.e., assessment for learning) and definitions that reflect nuanced differences among researchers. For example, some focus on the actions of the instructor (e.g., Ruiz-Primo and Furtak, 2006; Offerdahl and Tomanek, 2011;

Marilyne Stains *Monitoring Editor*

Submitted Mar 1, 2018; Revised Aug 29, 2018;

Accepted Sep 10, 2018

CBE Life Sci Educ December 1, 2018 17:es16

DOI:10.1187/cbe.18-02-0029

*Address correspondence to: Erika G. Offerdahl (erika.offerdahl@wsu.edu).

© 2018 E. G. Offerdahl *et al.* CBE—Life Sciences Education © 2018 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>). “ASCB®” and “The American Society for Cell Biology®” are registered trademarks of The American Society for Cell Biology.

Sadler and Reimann, 2018), others on the behaviors and predispositions of students (e.g., Gibbs and Simpson, 2004; Clark, 2012; Hepplestone and Chikwa, 2014). Differences notwithstanding, current definitions include representations of formative assessment as a process by which evidence of student learning is used to generate feedback to both the learner and instructor about progress toward desired learning outcomes (e.g., Sadler, 1998; Furtak and Ruiz-Primo, 2008; Bennett, 2011). Thus, formative assessment is thought to positively impact student learning through ongoing diagnosis and feedback about in-progress learning. We propose that, given the strong theoretical support for the efficacy of formative assessment, variation in effect sizes may reflect differences in enactment by instructors “in the wild” and not the efficacy of formative assessment as a high-impact instructional practice (Yin et al., 2008; Stains and Vickrey, 2017). Observed differences in outcomes between the theorized or “ideal” practice and enacted practice can be described in terms of fidelity of implementation (FOI) or “the extent to which an enacted program is consistent with the intended program model” (Century et al., 2010, p. 202).

We adopt an FOI perspective (Mowbray et al., 2003; Century et al., 2010; Stains and Vickrey, 2017) to generate a preliminary hypothesis about the critical components of formative assessment for learning and, in so doing, begin the iterative process through which further research can develop valid and reliable measures of the FOI of formative assessment. Such measures are necessary to empirically determine when, how, and under what conditions formative assessment supports student learning. Accordingly, we begin by defining formative assessment. We then present evidence in support of the hypothesized critical components and discuss potential mediating factors. Finally, we propose potential *adaptations* of the critical components and discuss the *productivity* of these adaptations in terms of student learning. We close with a brief discussion of the next steps for research within the context of student-centered undergraduate life sciences instruction. While the primary audience for this work is other discipline-based education researchers, when appropriate, we also include implications for practitioners.

DEFINING FORMATIVE ASSESSMENT

Historically, undergraduate teaching has been predominantly transmissionist in nature; instruction has been structured with the goal of conveying information to students (Barr and Tagg, 1995). Not surprisingly, within the context of these traditional learning environments, formative assessment has been employed as intermittent measurement of learning, an activity whereby the instructor conveys information to the student about the relative correctness, strengths or weaknesses, and appropriateness of the student’s work (Nicol and Macfarlane-Dick, 2006). In transmissionist instructional environments, the locus of control over assessment and feedback resides solely with the instructor. Opportunities for students to develop skills associated with self-regulated learning are therefore limited. Furthermore, while the delivery of feedback to students is accounted for, it largely neglects whether and how students translate feedback into actions that influence learning. As a result, formative assessment in these environments is difficult to distinguish from summative assessment.

With the growing adoption of evidence-based teaching practices and increasingly student-centered instruction, trans-

missionist instructional models are becoming less common. As a result, the paradigm of formative assessment has necessarily evolved in tandem from a unidirectional instructor monologue to an interactive multidirectional assessment conversation between and among students and members of the instructional team. Accordingly, a more contemporary definition depicts formative assessment as an iterative and reflective process or “dialogue” that creates a feedback loop through which members of the instructional team and students use evidence of student understanding to monitor progress toward learning outcomes and adapt practices to achieve those learning outcomes (Shepard, 2000; Bell and Cowie, 2001; Nicol and Macfarlane-Dick, 2006; Furtak and Ruiz-Primo, 2008).

This dialogue (Figure 1) most often begins with a *formative assessment prompt* (e.g., clicker question), an opportunity for students to engage with the material and test their understanding, thereby making their thinking explicit (e.g., Allen and Tanner, 2005; Knight and Wood, 2005). Ideally, *evidence of student understanding* informs both the learner and the instructor about students’ in-progress learning (Sadler, 1998; Furtak and Ruiz-Primo, 2008; Bennett, 2011). Diagnosis of learning progress can influence student and instructor behaviors. For example, productive *student behaviors* might include engaging in self-regulated learning processes, active reflection on his or her current level of understanding as compared with the desired learning outcome, and internally generated feedback (Nicol and Macfarlane-Dick, 2006), thereby prompting subsequent behaviors such as adopting alternative study strategies or seeking more information from additional external resources (e.g., textbook, peer, teaching assistant).

Evidence of student understanding can also inform *instructor behaviors*. If a gap between students’ current and desired level of understanding is perceived, the instructor might revise instruction or create new instructional activities (Hattie, 2012). This action could then initiate the start of a new cycle with another formative assessment prompt, with the goal of engaging students and eliciting evolution in student understanding. Alternatively, or additionally, the instructor can provide feedback to students about their in-progress learning. Ideally, this feedback incites further student action (e.g., engaging metacognitive processes, revising study strategies). We note here that, while the cycle frequently but not always begins with an instructor-generated formative assessment prompt, it is possible for an unsolicited student question to reveal evidence of student understanding.

In practice, formative assessment would likely be implemented as a cycle that would be repeated and spiraled forward, supporting students as they move toward the achievement of learning outcomes. Within student-centered learning environments, formative assessment is therefore conceived as both a measure of in-progress learning and a process to scaffold learning.

FOI: A FRAMEWORK FOR UNDERSTANDING THE IMPACT OF FORMATIVE ASSESSMENT

FOI, an approach rooted in the field of program evaluation, has recently been applied to research on the effectiveness of instructional practices in undergraduate science, technology, engineering, and mathematics (STEM). As applied to undergraduate STEM, an FOI perspective would assert that outcomes of

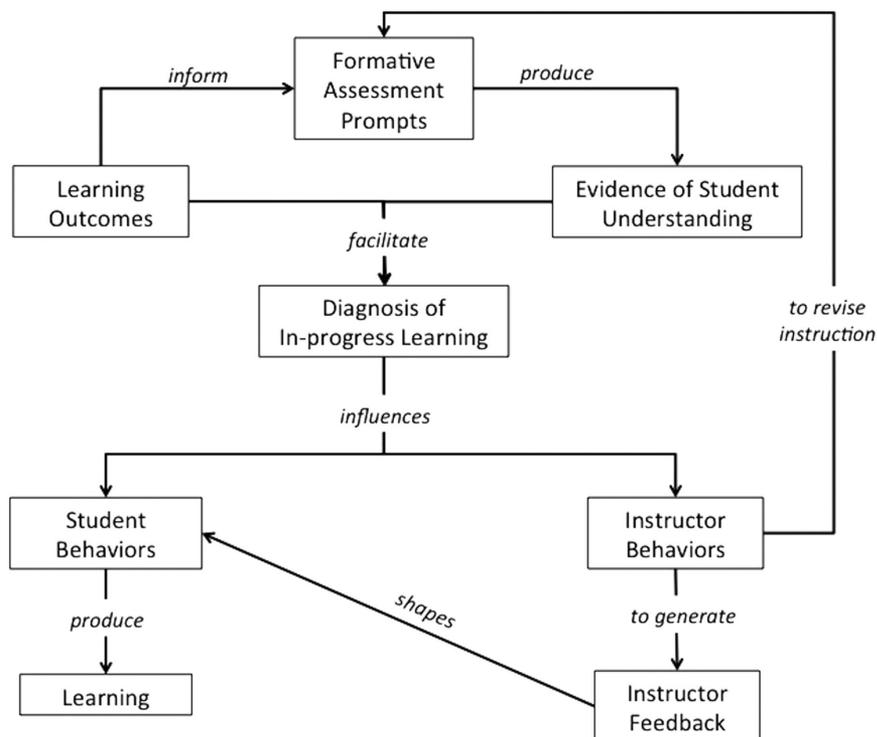


FIGURE 1. Formative assessment as an iterative process through which instructors and students use evidence of student understanding to monitor and generate actionable feedback to support progress toward desired learning outcomes.

common student-centered practices (e.g., peer instruction) are determined by the extent to which the *critical components* of an intended instructional practice are present during enactment (Stains and Vickrey, 2017). A critical component is defined as a feature or component that is essential to the instructional practice (Ruiz-Primo, 2006). Variations in how the critical components of an instructional practice are put into play are *adaptations*, the effects of which may be variable on the desired outcome (i.e., student learning). This variation can be explained, in part, by the myriad of *moderating variables* that are not critical components but still affect implementation and outcomes. While moderating variables by definition have a significant impact on implementation, they are not generally considered part of FOI itself (Ruiz-Primo, 2006).

Stains and Vickrey (2017) examined the potential impact of a well-known instructional practice, peer instruction (PI), through an FOI approach. PI is a student-centered pedagogical approach (Mazur, 1997) whereby students are asked to respond to a conceptual question that targets common student misconceptions. Typically, this response is provided using clickers or voting cards the students hold up. After they commit individually to an answer, students engage in discussion with their peers in an effort to convince them of the correct answer, and then vote individually again (Vickrey *et al.*, 2015). A critical component of PI is question difficulty—only about two-thirds of students should get the right answer on the first try. Another component is discussion with peers following the first vote and before the second vote (Stains and Vickrey, 2017).

In practice, there are a number of ways in which an instructor might implement the critical components that comprise PI.

A variation in the cognitive level of the question actually posed by an instructor (i.e., recall vs. application) represents an adaptation of that critical component (Stains and Vickrey, 2017). If an instructor were to implement PI by using predominantly recall questions, the FOI would be considered low (Century *et al.*, 2010) and might result in smaller learning gains than desired. Moreover, comparisons of outcomes between canonical PI and low-fidelity PI may result in misleading conclusions about the efficacy of the instructional practice.

For this critical component of PI, there is substantial empirical evidence to predict how productive an adaptation might be in terms of student learning; higher-level questions lead to greater learning gains than lower-order questions (e.g., Smith *et al.*, 2009; Knight *et al.*, 2013). Another critical component of PI is interspersing questions throughout lecture. Two possible adaptations could be 1) questions interspersed equally and 2) questions interspersed only in the second half of the lecture. Currently, there is little empirical work to determine the effect of these adaptations on student learning, and it may even be proven that no effect exists.

That is to say, adaptations of critical components may be productive, neutral, or unproductive for the desired outcome (Mills and Ragan, 2000; Ruiz-Primo, 2006; Century *et al.*, 2010).

The ways in which an instructor chooses to implement an instructional practice will determine how critical components are adapted. With respect to examining the implementation of instructional practices, potential moderating variables are readily identifiable in the literature. For example, research on the degree to which STEM instructors adopt evidence-based instructional practices (such as PI) are heavily influenced by personal and situational variables such as department reward structure, teaching context, and personal beliefs (Gess-Newsome *et al.*, 2003; Henderson and Dancy, 2007).

In the next section, we continue to build on the vast formative assessment literature base by adopting a “structure and process” approach (Century *et al.*, 2010) to operationalize formative assessment in terms of its critical components. Previous literature has sought to more clearly define formative assessment (e.g., Andrade and Cizek, 2010; Bennett, 2011), distinguish formative from summative assessment (e.g., Harlen and James, 1997; Bell and Cowie, 2001; Garrison and Ehringhaus, 2007), and expressly acknowledge the instructional and disciplinary contexts in which formative assessment is situated (e.g., Yorke, 2003; Coffey *et al.*, 2011). Many authors have taken a holistic approach to formative assessment by identifying overarching characteristics and/or strategies, but fewer have examined the relationships between characteristics (e.g., Bell and Cowie, 2001; Black and Wiliam, 2009). We propose that an FOI approach will expand on this work through a finer-grained, and in some ways mechanistic, look at the components of formative

TABLE 1. Criteria for characterizing level of support for proposed critical components

Level of support	Qualifier	Criterion
4	Strong support	Critical component is the 1) focus of a research question in two or more studies using different research methods or in different contexts and/or 2) a theoretical construct in more than two studies.
3	Moderate support	Critical component is the 1) focus of a single study and/or two or more studies with indirect evidence in support of the component and 2) a theoretical construct in more than two studies.
2	Limited support	Critical component is the 1) focus of studies with indirect evidence in support of the component and/or 2) a theoretical construct in at least two studies.
1	Theoretical support established	No more than two studies providing theoretical support, but no empirical studies.

assessment to answer questions about how different patterns of enactment impact student learning outcomes (Ruiz-Primo, 2006).

APPLICATION OF FOI TO IDENTIFY THE CRITICAL COMPONENTS OF FORMATIVE ASSESSMENT

Previous FOI work has applied three main approaches to identifying potential components: 1) leveraging empirical literature, 2) gathering input from experts, and 3) eliciting opinions of users and advocates about what works (Mowbray *et al.*, 2003; Stains and Vickrey, 2017). The first approach consists of identifying the quantifiable evidence of the proven efficacy and effectiveness of formative assessment by reviewing empirical studies and relevant meta-analyses. The second includes conducting expert surveys and/or literature reviews to articulate “key ingredients” that do not yet have operationalized criteria that can be verified empirically. The last is often approached via qualitative research methods that shed light on the questions of

“how” emerging from empirical studies. We used all three approaches in a recursive process. We began with a review of the literature and our own experiences as “users” of formative assessment to identify potential critical components. We then consulted the empirical and theoretical literature and applied similar criteria to Stains and Vickrey (2017) and Greenhalgh and colleagues (2004) to identify the level of support for each of the critical components (see Table 1). While leveraging the empirical literature, we reflected on the potential critical components in light of theoretical literature (and vice versa), our own experiences, and the experiences of other experienced instructors to identify potential critical components that also have strong empirical and/or theoretical support (see Table 2 and Supplemental Table S1).

The potential critical components can be further categorized into two groups, structural and instructional (Century *et al.*, 2010). The *structural* group refers to those components that reflect the composition and organization of an intervention. For

TABLE 2. Proposed critical components of formative assessment for which there is strong (empirical or theoretical) support as defined in Table 1

Category	Critical component	Description
Structural	Learning objectives	Clear criteria for success are identified.
	Formative assessment prompts	Mechanisms for eliciting the range and extent of students’ understanding are employed.
	Evidence of student understanding	Range and extent of student understanding is made explicit to teacher and student.
	Feedback	A comparison of the learner’s current state with the criteria for success is used to generate timely, relevant, and actionable feedback.
	Skills for self-regulated learning	Students know how to identify personal strengths/weaknesses relevant to the instructional task and how to create and monitor a plan for completing a learning task.
	Personal pedagogical content knowledge (PCK)	Instructors possess discipline-specific and pedagogical knowledge for designing and reflecting on instruction of particular topics.
	Prior knowledge	Students’ prior knowledge is activated and interacts with how they learn information.
Instructional	Reveal student understanding	The student/class willingly provides an appropriate response to the formative assessment prompt.
	Personal pedagogical knowledge and skills (PCK&S)	The instructor uses particular discipline-specific knowledge and pedagogical skills to diagnose learning of a particular topic and provide feedback in a particular way to particular students.
	Diagnosis of in-progress learning	The instructor and/or student uses formative assessment prompt and learning outcome to diagnose learner’s current state.
	Generate feedback	The instructor and/or student generate(s) feedback about the learner’s current state.
	Recognize and respond to feedback	The student recognizes and acts on feedback to shape learning.

example, structural critical components might include the prescribed instructional materials and the order in which they are presented to students. Structural components would also include the requisite knowledge needed by an instructor to enact the materials as intended. The *instructional* (process) group includes components related to the expected roles and/or behaviors—the human interactions—that take place during the intervention. Century *et al.* (2010) refer to the process group as instructional critical components, which reflect how instructors and students behave and interact with one another.

While the level of empirical evidence for each of the proposed components varies, each one has substantial theoretical support. The proposed critical components in Table 2 should be viewed as a starting point for measuring the FOI of formative assessment. The FOI literature indicates that initial sets of critical components will be iteratively refined as researchers 1) identify productive and unproductive variations in enactment (adaptations) of each critical component and 2) develop valid and reliable measures of each component (Mills and Ragan, 2000; Ruiz-Primo, 2006). Accordingly, in the next section, we lay the groundwork for future research by predicting potential *adaptations*—variations in the enactment of one or more critical components—and discuss the relative productivity of these adaptations as they relate to student learning.

POTENTIAL ADAPTATIONS OF THE CRITICAL COMPONENTS OF FORMATIVE ASSESSMENT

FOI has been defined as “the extent to which the critical components of an intended educational program, curriculum, or instructional practice are present when that program, curriculum, or practice is enacted” (Stains and Vickrey, 2017, p. 2). Adaptations in the implementation of each component may positively or negatively impact the predicted outcome of the intended instructional practice or may have no effect at all. We used our own teaching experience, prior research, and the literature to hypothesize about likely adaptations of the critical components. While by no means an exhaustive list, we dedicate the remainder of this section to exploring the productivity of four potential adaptations.

Adaptation 1: Variation in Alignment between Learning Outcomes and Formative Assessment Prompt

The power of formative assessment is in its ability to diagnose students’ current level of understanding relative to the desired level (targeted learning outcome). Without a clear indication of the desired learning outcome—what students need to know or be able to do—it is impossible to diagnose student progress, let alone revise instruction or generate meaningful feedback to improve student learning. It follows, then, that effective formative assessment prompts should first and foremost align with the stated learning outcomes; they must uncover students’ current understanding relative to the desired level of understanding (Wiggins and McTighe, 2005; Handelsman *et al.*, 2007; President’s Council of Advisors on Science and Technology [PCAST], 2012).

Consider the desired learning outcome for students to generate a scientific argument. Asking students to write a “Muddiest point” paper about what they do not understand at the end of class is unlikely to provide evidence about their argumentation skills and represents a potentially unproductive adaptation.

On the other hand, a productive adaptation would be modifying a minute-paper approach to ask students to generate and justify one claim from a graph. A minute paper would potentially reveal students’ abilities to interpret data and generate claims, both of which are essential to scientific argumentation. In these two examples, a formative assessment prompt was expressly implemented in response to a desired learning outcome, but the degree of alignment influenced the productivity of the implementation in terms of student learning. Misaligned outcomes and prompts are not likely to produce information that allows for instructors or students to diagnose potential gaps between the present and desired levels of student understanding.

Adaptation 2: Variation in Evidence of Student Understanding Produced by Formative Assessment Prompts

Diagnosis of students’ progress toward intended learning outcomes is dependent on sufficient information about the current state of students’ understanding or abilities. Instructors need to identify not just what students know, but importantly, what students are struggling with, and how prevalent those difficulties are (Furtak and Ruiz-Primo, 2008; Offerdahl and Montplaisir, 2014). In practice, this may be challenging. Large class sizes common to introductory undergraduate biology courses often favor the use of closed-ended prompts, such as forced-choice questions (e.g., clicker questions), due to their ease of application and their capacity to reveal the pervasiveness of ideas within the class (Smith and Tanner, 2010). Clicker questions in particular have the added advantage of creating an opportunity for students to immediately compare their answers with others and receive feedback about the relative correctness of those responses (Preszler *et al.*, 2007). Therefore, the use of clickers to elicit evidence of student understanding is a potentially productive adaptation. But the response options are typically instructor-generated, thereby only providing insights into student understanding about those specific topics selected by the instructor. Closed-ended prompts may fail to reveal unanticipated student ideas not previously documented in the literature or through prior teaching experiences (Offerdahl and Montplaisir, 2014). Use of such prompts could potentially be unproductive for student learning when uncovering diversity of student thinking is appropriate.

Open-ended formative assessment prompts provide a different opportunity for examining students’ understanding. For example, concept maps or student-generated models invite students to represent how they personally make connections between ideas (e.g., Mintzes *et al.*, 2005). The potential merit of open-ended prompts is that they provide an opportunity to reveal the diversity with which students approach and make sense of phenomena and identify overarching trends within the class. On the other hand, if the prompt is too open-ended (i.e., muddiest point) students may provide information that is not directly related to the stated learning outcome or may provide little information—students don’t always know what they don’t know (Tanner and Allen, 2004; Offerdahl and Montplaisir, 2014). Alternatively, the prompt may generate an overwhelmingly large range of student ideas that are difficult for an instructor to summarize and respond to concisely.

Evidence of student understanding is a critical component of formative assessment; the cycle will not work without

diagnosing what students know and do not know in relation to the desired learning outcome. Yet there is little empirical evidence to explain how variations in this critical component ultimately influence student learning.

Adaptation 3: Variation in Instructors' Pedagogical Content Knowledge

We did not explicitly depict pedagogical content knowledge (PCK) in the model (Figure 1), though the literature has often drawn connections between PCK and instructional practice (e.g., Shulman, 1986; Abell, 2008; Etkina, 2010; Van Driel and Berry, 2012). Recently, PCK has been more concisely defined to encompass two related constructs, personal PCK and personal PCK and skills (PCK&S). Personal PCK is an instructor's "knowledge of, reasoning behind, and planning for teaching a particular topic in a particular way for a particular purpose to particular students" (Gess-Newsome, 2015, p. 36). It can be thought of as the disciplinary and pedagogical knowledge drawn on by an instructor when planning for teaching. As such, it can be characterized as a structural critical component much like other types of instructor knowledge. PCK&S is a closely related construct; it involves activation of personal PCK while engaging in the act of teaching a specific topic in a certain way to achieve a particular outcome. PCK&S is likely an *instructional* critical component, in that it involves the real-time actions of instructors in response to the present context.

It has been suggested that teachers may need a basic level of personal PCK to enact an instructional practice; yet some teachers may come to the classroom with that knowledge, while others may not (Century *et al.*, 2010). Biology instructors with well-developed personal PCK are readily able to articulate the "big ideas" in biology due to their content expertise and possess the pedagogical skills necessary to generate measurable learning outcomes that reflect what a student should know or be able to do relevant to those big ideas. Similarly, a combination of both content expertise and pedagogical know-how is required for instructors to create formative assessment prompts that not only align with learning outcomes but are likely to produce rich evidence of student understanding, including common reasoning difficulties specific to core ideas in biology.

Both personal PCK and PCK&S are necessary for an instructor to make sense of, respond to, and use evidence of student understanding effectively in real time. For example, responding to students' misconceptions about the role of noncovalent interactions in protein structure requires knowledge of common student reasoning difficulties about atomic structure, bonding, and polarity (personal PCK) and creating meaningful opportunities in class for students to restructure their mental models to reflect more scientifically accurate understanding of the concept (PCK&S). Similarly, an instructor must have knowledge about effective feedback (i.e., relevant, timely, actionable) and the skills to provide that feedback within the context of a particular topic to guide students in achieving particular outcomes. In each of these examples, well-developed personal PCK and PCK&S are likely a productive adaptation.

K–12 teacher preparation programs have long recognized the importance of instructors' PCK for effective science instruction (Abell, 2008; Van Driel and Berry, 2012), and further teacher professional development is associated with growth in personal PCK (Daehler *et al.*, 2015). Perhaps unsurprisingly,

professional development for life sciences' instructors similarly incorporates opportunities to learn about and apply personal PCK (e.g., National Academies Summer Institutes on Undergraduate Education, Faculty Institutes for Reforming Science Teaching).

Adaptation 4: Variation in Actionable Instructor-Generated Feedback

Feedback is a proposed structural critical component that links evidence of student understanding to students' learning activities (Figure 1). In a meta-analysis of feedback interventions, Kluger and DeNisi (1996) detected an average positive effect ($d = 0.41$) of feedback on student learning, yet in more than 38% of the studies, a negative effect was measured. These data suggest that there are potentially productive and unproductive adaptations of feedback. Providing timely, relevant, and actionable feedback to students on their learning is likely to be productive for student learning, particularly if that feedback provides answers to the questions of "Where am I going?," "How am I doing?," and "Where to next?" Such feedback provides clear guidance for students on their learning (Hattie and Timperley, 2007).

It is intuitive to propose that the converse condition, feedback that is not timely or relevant, is likely an unproductive adaptation. Students are more likely to act on feedback that is relevant (Hepplestone and Chikwa, 2014) and promotes a dialogue and relationship with instructors (Nicol, 2010; Price *et al.*, 2010). Moreover, there are other potentially unproductive adaptations, such as nonspecific praise or simple verification of whether the answer is correct, that have been shown to negatively effect student learning (Hattie and Timperley, 2007).

Implications of an FOI Approach for Undergraduate Biology Instructors and Biology Education Researchers

In the preceding sections, we used an FOI approach to 1) identify potential critical components of formative assessment that must be present to positively impact student learning and 2) describe adaptations of the critical components and their impact on learning. We argue that such an approach provides a useful lens for interpreting the effect sizes associated with formative assessment previously reported in the literature; namely, differences may be attributed to adaptations of the critical components of formative assessment rather than the efficacy of formative assessment as an instructional practice itself. In this section, we briefly discuss implications of the application of an FOI perspective for practitioners and for future research in undergraduate biology.

Current education reform documents call for undergraduate instructors to implement student-centered instructional practices with frequent and ongoing assessment of students' in-progress learning, and to use assessment data to improve learning (American Association for the Advancement of Science, 2011; PCAST, 2012). Reflecting on one's teaching has long been recognized as a process through which instructors can turn personal experience into knowledge about teaching that can support further development of the instructor and associated practices (e.g., McAlpine *et al.*, 1999; Kreber, 2005; Andrews and Lemons, 2015). But to maximize the utility of reflection, instructors must know what cues to monitor and how to evaluate them (McAlpine *et al.*, 1999; McAlpine and Weston, 2000).

TABLE 3. Nonexhaustive list of existing tools that could potentially be used and/or modified to measure aspects of FOI of formative assessment

Measurement tools	Relevant critical component(s)	References
Bloom's taxonomy tools	Learning outcomes, formative assessment prompts	Crowe <i>et al.</i> , 2008; Arneson and Offerdahl, 2018
ESRU cycle ^a	Formative assessment prompts, evidence of student understanding, feedback	Ruiz-Primo and Furtak, 2007; Furtak <i>et al.</i> , 2017
ICAP framework ^b	Formative assessment prompts	Chi and Wylie, 2014
Project PRIME PCK rubric ^c	Personal pedagogical content knowledge	Gardner and Gess-Newsome, 2011
Practical Observation Rubric To Assess Active Learning (PORTAAL)	Learning outcomes, formative assessment prompts, feedback	Eddy <i>et al.</i> , 2015
Concept inventories (e.g., EcoEvo-Maps)	Prior knowledge	Summers <i>et al.</i> , 2018
Metacognitive Awareness Inventory	Skills for self-regulated learning	Harrison and Vallin, 2018

^aThe ESRU cycle is characterized by an instructor Eliciting evidence of Student thinking followed by an instructor Recognizing the response and Using it to support learning.

^bThe ICAP framework uses student behaviors to characterize cognitive engagement activities as interactive, constructive, active, and/or passive.

^cProject PRIME (Promoting Reform through Instructional Materials that Educate) produced the PCK rubric with support from the National Science Foundation (DRL-0455846).

Explicit identification of the critical components of formative assessment makes explicit the potential cues that should be monitored. Similarly, FOI emphasizes noticing the various adaptations in implementation of critical components, and the potential effects on the desired outcome. This approach facilitates systematic reflection on an instructor's personal formative assessment practice as it relates to student learning. For example, there is strong support that formative assessment prompts that elicit the range and extent of student understanding are a critical component of effective formative assessment. With this knowledge, an instructor can reflect on the degree to which a formative assessment prompt is providing sufficient evidence of student understanding and adjust if necessary. FOI is useful approach for instructors to encourage systematic reflection on enactment of formative assessment and to make sense of the implications of those actions.

For biology education researchers, FOI is a useful approach for designing studies that will lead to even more valid claims about the effectiveness of particular student-centered instructional practices in diverse settings. This is especially important at a time when the biology education research community has moved beyond comparison studies between traditional and active, student-centered instruction (Dolan, 2015). The evidence clearly indicates that active, student-centered learning produces greater learning gains (Prince, 2004; Freeman *et al.*, 2014). As demonstrated here, FOI is an approach for researchers to systematically propose, refine, and measure the effects of the "key ingredients" of various student-centered instructional practices, thereby facilitating comparison between contexts. With respect to formative assessment in particular, this approach facilitates leveraging and/or refining existing quantitative and qualitative tools (Table 3) and developing further measures of the critical components of formative assessment. In doing so, researchers will be better equipped to tease apart reported differences in outcomes associated with various student-centered practices.

CONCLUSIONS

We have applied an FOI perspective to the potentially high-impact instructional practice of formative assessment as a first step in laying the foundation for more empirical work investigating when, how, and under what conditions formative

assessment supports student learning. Consistent with previous FOI work, future work should strive to 1) iteratively refine the critical components identified here, 2) establish valid and reliable measures of each critical component, and 3) design efficacy and effectiveness studies using an FOI approach. Such work will allow the biology education research community to more effectively document how variations in implementation of formative assessment impact learning in modern, student-centered undergraduate biology contexts.

ACKNOWLEDGMENTS

We thank Andy Cavagnetto, Jenny Momsen, Aramati Casper, Jessie Arneson, and members of the PULSE community for insightful comments on previous versions of this essay. We thank the monitoring editor and external reviewers for their time and constructive comments. This material is based on work supported by the National Science Foundation under grant no. 1431891.

REFERENCES

- Abell, S. K. (2008). Twenty years later: Does pedagogical content knowledge remain a useful idea?. *International Journal of Science Education*, 30(10), 1405–1416.
- Allen, D., & Tanner, K. (2005). Infusing active learning into the large-enrollment biology class: Seven strategies, from the simple to complex. *Cell Biology Education*, 4(4), 262–268.
- American Association for the Advancement of Science. (2011). *Vision and change in undergraduate biology education: A call to action: A summary of recommendations made at a national conference organized by the American Association for the Advancement of Science, July 15–17, 2009*. Washington, DC.
- Andrade, H., & Cizek, G. J. (2010). An introduction to formative assessment: History, characteristics, and challenges. In Andrade, H., & Cizek, G. J. (Eds.), *Handbook of formative assessment* (pp. 15–29). New York: Routledge.
- Andrews, T. C., & Lemons, P. P. (2015). It's personal: Biology instructors prioritize personal evidence over empirical evidence in teaching decisions. *CBE—Life Sciences Education*, 14(1), ar7.
- Arneson, J. B., & Offerdahl, E. G. (2018). Visual literacy in Bloom: Using Bloom's taxonomy to support visual learning skills. *CBE—Life Sciences Education*, 17(1), ar7.
- Barr, R. B., & Tagg, J. (1995). From teaching to learning—A new paradigm for undergraduate education. *Change: The Magazine of Higher Learning*, 27(6), 12–26.

- Bell, B., & Cowie, B. (2001). The characteristics of formative assessment in science education. *Science Education*, 85(5), 536–553.
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5–25.
- Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74.
- Black, P., & Wiliam, D. (1998b). Inside the black box. *Phi Delta Kappan*, 80(2), 139–148.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5.
- Briggs, D. C., Ruiz-Primo, M. A., Furtak, E., Shepard, L., & Yin, Y. (2012). Meta-analytic methodology and inferences about the efficacy of formative assessment. *Educational Measurement: Issues and Practice*, 31(4), 13–17.
- Century, J., Rudnick, M., & Freeman, C. (2010). A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. *American Journal of Evaluation*, 31(2), 199–218.
- Chi, M. T., & Wylie, R. (2014). The ICAP framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219–243.
- Clark, I. (2012). Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review*, 24(2), 205–249.
- Coffey, J. E., Hammer, D., Levin, D. M., & Grant, T. (2011). The missing disciplinary substance of formative assessment. *Journal of Research in Science Teaching*, 48(10), 1109–1136.
- Crowe, A., Dirks, C., & Wenderoth, M. P. (2008). Biology in Bloom: Implementing Bloom's taxonomy to enhance student learning in biology. *CBE—Life Sciences Education*, 7(4), 368–381.
- Daehler, K. R., Heller, J. I., & Wong, N. (2015). Supporting growth of pedagogical content knowledge in science. In Berry, A., Friedrichse, P., & Loughran, J. (Eds.), *Re-examining pedagogical content knowledge in science education* (pp. 45–59). New York: Routledge.
- Dolan, E. L. (2015). Biology education research 2.0. *CBE—Life Sciences Education*, 14, ed1.
- Dunn, K. E., & Mulvenon, S. W. (2009). A critical review of research on formative assessment: The limited scientific evidence of the impact of formative assessment in education. *Practical Assessment, Research & Evaluation*, 14(7), 1–11.
- Eddy, S. L., Converse, M., & Wenderoth, M. P. (2015). PORTAAL: A classroom observation tool assessing evidence-based teaching practices for active learning in large science, technology, engineering, and mathematics classes. *CBE—Life Sciences Education*, 14(2), ar23.
- Etkina, E. (2010). Pedagogical content knowledge and preparation of high school physics teachers. *Physical Review Special Topics—Physics Education Research*, 6(2), 020110.
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences USA*, 111(23), 8410–8415.
- Furtak, E. M., & Ruiz-Primo, M. A. (2008). Making students' thinking explicit in writing and discussion: An analysis of formative assessment prompts. *Science Education*, 92(5), 799–824.
- Furtak, E. M., Ruiz-Primo, M. A., & Bakeman, R. (2017). Exploring the utility of sequential analysis in studying informal formative assessment practices. *Educational Measurement: Issues and Practice*, 36(1), 28–38.
- Gardner, A. L., & Gess-Newsome, J. (2011, April). A PCK rubric to measure teachers' knowledge of inquiry-based instruction using three data sources. Paper presented at: *Annual Meeting of the National Association for Research in Science Teaching, Orlando, FL*. Retrieved May 22, 2018, from www.bscs.org/sites/default/files/_legacy/pdf/Community_Sessions_NARST2011_PCK%20Rubric%20Paper.pdf
- Garrison, C., & Ehringhaus, M. (2007). Formative and summative assessments in the classroom. Retrieved February 23, 2018, from www.amle.org/Publications/WebExclusive/Assessment/tabid/1120/Default.aspx
- Gess-Newsome, J. (2015). A model of teacher professional knowledge and skill including PCK: Results of the thinking from the PCK Summit. In Berry, A., Friedrichse, P., & Loughran, J. (Eds.), *Re-examining pedagogical content knowledge in science education* (pp. 22–42). New York: Routledge.
- Gess-Newsome, J., Southerland, S. A., Johnston, A., & Woodbury, S. (2003). Educational reform, personal practical theories, and dissatisfaction: The anatomy of change in college science teaching. *American Educational Research Journal*, 40(3), 731–767.
- Gibbs, G., & Simpson, C. (2004). Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education* 82, 3–31.
- Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., & Kyriakidou, O. (2004). Diffusion of innovations in service organizations: Systematic review and recommendations. *Milbank Quarterly*, 82(4), 581–629.
- Handelsman, J., Miller, S., & Pfund, C. (2007). *Scientific teaching*. New York: W. H. Freeman.
- Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy & Practice*, 4(3), 365–379.
- Harrison, G. M., & Vallin, L. M. (2018). Evaluating the metacognitive awareness inventory using empirical factor-structure evidence. *Metacognition and Learning*, 13(1), 15–38.
- Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. London, England: Routledge.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81–112.
- Henderson, C., & Dancy, M. H. (2007). Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics. *Physical Review Special Topics—Physics Education Research*, 3(2), 020102.
- Hepplestone, S., & Chikwa, G. (2014). Understanding how students process and use feedback to support their learning. *Practitioner Research in Higher Education*, 8(1), 41–53.
- Kingston, N., & Nash, B. (2011). Formative assessment: A meta-analysis and a call for research. *Educational Measurement: Issues and Practice*, 30(4), 28–37.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254–284.
- Knight, J. K., Wise, S. B., & Southard, K. M. (2013). Understanding clicker discussions: Student reasoning and the impact of instructional cues. *CBE—Life Sciences Education*, 12(4), 645–654.
- Knight, J. K., & Wood, W. B. (2005). Teaching more by lecturing less. *Cell Biology Education*, 4(4), 298–310.
- Kreber, C. (2005). Reflection on teaching and the scholarship of teaching: Focus on science instructors. *Higher Education*, 50(2), 323–359.
- Mazur, E. (1997). Peer instruction: Getting students to think in class. *AIP Conference Proceedings* 399(1), 981–988.
- McAlpine, L., & Weston, C. (2000). Reflection: Issues related to improving professors' teaching and students' learning. *Instructional Science*, 28(5), 363–385.
- McAlpine, L., Weston, C., Beauchamp, C., Wiseman, C., & Beauchamp, J. (1999). Building a metacognitive model of reflection. *Higher Education*, 37(2), 105–131.
- Mills, S. C., & Ragan, T. J. (2000). A tool for analyzing implementation fidelity of an integrated learning system. *Educational Technology Research and Development*, 48(4), 21–41.
- Mintzes, J. J., Wandersee, J. H., & Novak, J. D. (2005). *Assessing science understanding: A human constructivist view*. San Diego, CA: Academic Press.
- Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation*, 24(3), 315–340.
- Nicol, D. (2010). From monologue to dialogue: Improving written feedback processes in mass higher education. *Assessment & Evaluation in Higher Education*, 35(5), 501–517.
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218.
- Offerdahl, E. G., & Montplaisir, L. (2014). Student-generated reading questions: Diagnosing student thinking with diverse formative assessments. *Biochemistry and Molecular Biology Education*, 42(1), 29–38.

- Offerdahl, E. G., & Tomanek, D. (2011). Changes in instructors' assessment thinking related to experimentation with new strategies. *Assessment & Evaluation in Higher Education, 36*(7), 781–795.
- President's Council of Advisors on Science and Technology. (2012). *Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics*. Washington, DC: U.S. Government Office of Science and Technology.
- Preszler, R. W., Dawe, A., Shuster, C. B., & Shuster, M. (2007). Assessment of the effects of student response systems on student learning and attitudes over a broad range of biology courses. *CBE—Life Sciences Education, 6*(1), 29–41.
- Price, M., Handley, K., Millar, J., & O'Donovan, B. (2010). Feedback: All that effort, but what is the effect?. *Assessment & Evaluation in Higher Education, 35*(3), 277–289.
- Prince, M. (2004). Does active learning work? A review of the research. *Journal of Engineering Education, 93*(3), 223–231.
- Ruiz-Primo, M. A. (2006). *A multi-method and multi-source approach for studying fidelity of implementation*. (CSE Report 677). Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.
- Ruiz-Primo, M., & Furtak, E. M. (2006). Informal formative assessment and scientific inquiry: Exploring teachers' practices and student learning. *Educational Assessment, 11*(3–4), 237–263.
- Ruiz-Primo, M. A., & Furtak, E. M. (2007). Exploring teachers' informal formative assessment practices and students' understanding in the context of scientific inquiry. *Journal of Research in Science Teaching, 44*(1), 57–84.
- Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119–144.
- Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education: Principles, Policy & Practice, 5*(1), 77–84.
- Sadler, I., & Reimann, N. (2018). Variation in the development of teachers' understandings of assessment and their assessment practices in higher education. *Higher Education Research & Development, 37*(1), 131–144.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4–14.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher, 15*(2), 4–14.
- Smith, J. I., & Tanner, K. (2010). The problem of revealing how students think: Concept inventories and beyond. *CBE—Life Sciences Education, 9*(1), 1–5.
- Smith, M. K., Wood, W. B., Adams, W. K., Wieman, C., Knight, J. K., Guild, N., & Su, T. T. (2009). Why peer discussion improves student performance on in-class concept questions. *Science, 323*(5910), 122–124.
- Stains, M., & Vickrey, T. (2017). Fidelity of implementation: An overlooked yet critical construct to establish effectiveness of evidence-based instructional practices. *CBE—Life Sciences Education, 16*(1), rm1.
- Summers, M. M., Couch, B. A., Knight, J. K., Brownell, S. E., Crowe, A. J., Semsar, K., ... Smith, M. K. (2018). EcoEvo-MAPS: An ecology and evolution assessment for introductory through advanced undergraduates. *CBE—Life Sciences Education, 17*(2), ar18.
- Tanner, K., & Allen, D. (2004). Approaches to biology teaching and learning: From essays to assessments—on collecting evidence in science teaching. *Cell Biology Education, 3*(2), 69–74.
- Van Driel, J. H., & Berry, A. (2012). Teacher professional development focusing on pedagogical content knowledge. *Educational Researcher, 41*(1), 26–28.
- Vickrey, T., Rosploch, K., Rahmanian, R., Pilarz, M., & Stains, M. (2015). based implementation of peer instruction: A literature review. *CBE—Life Sciences Education, 14*(1), es3.
- Wiggins, G. P., & McTighe, J. (2005). *Understanding by design*. Alexandria, VA: ASCD.
- Yin, Y., Shavelson, R. J., Ayala, C. C., Ruiz-Primo, M. A., Brandon, P. R., Furtak, E. M., ... Young, D. B. (2008). On the impact of formative assessment on student motivation, achievement, and conceptual change. *Applied Measurement in Education, 21*(4), 335–359.
- Yorke, M. (2003). Formative assessment in higher education: Moves towards theory and the enhancement of pedagogic practice. *Higher Education, 45*(4), 477–501.